

Bibliometric Data as a Tool for the Analysis of the Scientist's Mobility

Anastasia E. Sudakova^{1, a)}, Alexander A. Tarasyev^{1, b)}

¹*Ural Federal University, 620002, Mira Street, 19, Yekaterinburg, Russia*

^{a)}Corresponding author: a-chusova@mail.ru

^{b)}a.a.tarasyev@urfu.ru

Abstract. The article covers the bibliometric approach to the analysis of migration of scientists and provides an overview of authors using this approach. In most cases, bibliometric data is accumulated from SCOPUS and Web of Science publications. In some cases, the authors used patents for data collection and data from the PubMed platform. The affiliate database cleansing is performed using CV (which is a separate tool for collecting bibliometric data in some cases). The database cleansing according to verified data (city, initials, and place of work) is performed, too. Identifiers, such as ResearcherID and ORCID, do not currently simplify the data processing.

INTRODUCTION

Scientific and technological development of the country is a priority of most governments. In our opinion, the basis of this development is the interconnection of the following elements - intellectual potential of the nation, including the level of education of the whole country and the formation of the scientific elite; the state policy forming the conditions for scientific activity.

Russia is in the top 50 of the innovative development index (ranking from 43 to 49 in 2019-2014), and in the top 40 in terms of domestic expenditure on R&D as a percentage of GDP (placed 35 in 2016). Equally negative dynamics develops regarding the intellectual capital of our country. A number of problems can be identified: emigration of scientists, low mobility and immigration of leading scientists.

The intellectual potential of the nation is a research task of our article. The study presents the research problems of quantifying the migration of scientists and assessing the impact of migration on the level of scientific effectiveness. In this article, we analyze the bibliometric approach to the analysis of the migration of scientists.

THEORETICAL FRAMEWORK

In order to learn how the national science works and what its strengths and weaknesses are, it is important to know about the cross-border flows of scientists, and it's equally important to know how mobility is structured nationwide.

The main difficulty in conducting this kind of research is the lack of reliable and complete information on the migration of scientists.

Official Russian statistics provides information only on the number of officially migrated citizens, i.e. recorded cases of migration. Such data do not allow us to analyze the migration of scientists. Firstly, the data is not divided into any categories. Secondly, the migration of scientists is not easy to assess: it can be short or long trips (mobility). In addition, it can be assumed that in modern times the decision on migration does not lead to a complete "legal" loss of connection with the country of departure (for example, quitting a job, deregistration, sale of all property). Consequently, we note that the migration of scientists has several record features: how to record the loss of

intellectual capital (through dual affiliation or at its complete change) and how to assess the impact of migration and its types (mobility) on the dissemination of knowledge and scientific effectiveness.

Data on mobility largely comes from census data, registration data, labor force surveys (e.g., CLFS), longitudinal panels (e.g., NCSSES, EuroSTAT, ECHP), individual and organization surveys, and specific case studies – none of which have been seen to be sufficient in providing comprehensive and contemporary analysis of scientific migration for policy purposes [1], [2], [3], [4].

Traditionally, mobility of scientist is studied using such methods as government statistics on academic staff and data on migration [5], [6], analysis of CVs and personal web pages [4], [7], [8], questionnaires and interviews with scientists themselves [9], [10], and government and administrative databases [11]. However, for the study of migration, including mobility of researchers used tools: social network, bibliometric data from the SCOPUS, web of science and others, among the works devoted to the collection of bibliometric data it is possible to note Du bois P. [12], Moed H. F. [13] and other authors, some of them are presented next.

This method makes it possible to carry out a comparative study of the publishing activity of mobile and non-mobile authors [14], and to study the impact of migration on the development of various disciplines [15]. By using the affiliations method, it is also possible to study the mobility of groups of elite scientists that are small, but nonetheless important for the development of science [16]. These 'digital traces' can capture the movement of scientists between countries [17]; the concentration of representatives of various disciplines in certain countries or organizations [18]; and enable analysis of relative migration flows [13].

Given that publications are associated with date information, we can conduct diachronic network analyses to identify the trade of scholars not only between locations, but among all locations over time. Publication data provide the added advantage that we can examine the impact of mobility, by measuring citations before, after, and during periods of transition. Furthermore, bibliometric data can be analyzed at least quarterly, which addresses the problem of the delays in obtaining statistics on R&D personnel that has been repeatedly noted in the literature [3].

Laudel [19] was one of the first to argue for the use of bibliometrics to construct global indicators on scientific mobility. Laudel [20] employed PubMed data, which provides first-author affiliations since 1980. She augmented her analysis with data on affiliation of doctoral degree. However, she restricted her analysis to a certain classification of elite scholars: that is, those who published at least three papers in Science and Nature between 1980 and 2002. This is a restrictive dataset, both in terms of disciplinary coverage as well as country coverage. Taking only first-author data suppresses the contributions of the scientists who labor in large collaborative teams. Looking at those who have published three times in elite journals further reduces the pool of eligible scientists. This is a persistent problem in the literature on mobility, with many studies focusing on the elite and super elite – for example, Nobel Prize winners [21], [22].

The research of Moed and Halevi [13], who examined migration balances between a select group of developing and developed countries, using Scopus data. This work was methodologically useful in that it both discussed and examined the difficulties of author-name disambiguation for bibliometric data, including complexities of homonyms and synonyms in the database. In their validation study of 100 randomly selected Chemistry authors, they conclude that database errors contribute to relatively little change in the results.

BIOMETRIC DATA PROCESSING

In this paragraph we will consider how the processing of bibliometric data. First, data processing includes validation of affiliations, elimination of double counting. Secondly, after a "clean" or "working" database is obtained, the next stage arises – how to build a model or network.

The issue of database verification is considered in the work of 2009 by Ekaterina L. Dyachenko [23] the processing was carried out by removing the authors whose articles were not found in the compared year, and authors for whom WoS contains too many papers, in the sense that it contains data on several authors with the same name who cannot be distinguished. In addition, the processing was carried out by using CVs as an additional source of data for distinguishing between authors with the same names. At the same time, the authors were initially selected randomly from the selected target category. The resulting network was visualized and analyzed by means of UCINET software (UCINET 6 for Windows).

Trajtenberg M. said that matching inventors purely on their names introduces the risk of false negative errors (inventors may use multiple spelling permutations of their name such that we miss actual movers) and false positive errors (different inventors may have the same name such that we flag someone as a mover who is not) [24]. Given that there are over 2 million patents with 2 inventors per patent on average, the "who is who" problem applies to

over 4 million "records", which is obviously too large to tackle manually. They have thus developed an elaborate methodology and computerized procedure to address this problem in a comprehensive way. The end result is a list of 1.6 million unique inventors from all over the world, with detailed data on their patenting histories, their employers, co-inventors, etc.

Given the possible errors in the identification of classified Trajtenberg, Agrawal A. [25] treats them as follows: false negative errors Agrawal A. claims that they are at an acceptable level of error, To minimize false positive errors, Agrawal A. add the sampling restriction that the inventor's pre-1990 patent must be in a similar technology field as their 1990 patent. Therefore, they impose the restriction that the 1990 patent's major three-digit classification must either match the prior patent's own major three-digit classification or be included in the prior patent's set of three-digit cross-classifications. Thus, they eliminate the problem of falsely identifying two inventors with the same name who are working in different fields and in different locations as the same person and hence a mover. However, it is important to note that Agrawal A. may still experience measurement error in our process for identifying movers. To the extent that they falsely identify movers (two or more individuals who have the same name and patent in the same field but live in different cities), our findings will be biased against their hypothesis.

Several unobtrusive methods have been used to derive mobility data for scientists: e.g., encyclopedias, biographical information, and curriculum vitae [19]. The use of these data, however, is inherently limited in scope, coverage, and accuracy and requires a substantial amount of cleaning, reducing the utility of the datasets for large-scale analyses (see, e.g., Dietz et al. [26], [27]). Laudel was one of the first to argue for the use of bibliometrics to construct global indicators on scientific mobility. Laudel [28] employed PubMed data

Among earlier works, at the time when ResearcherID and ORCID were just starting to form, it was believed that these identifiers would simplify the process of collecting bibliometric data. However, already on the basis of processing affiliations for publications of 2010-2015 presented in SCOPUS, it was concluded that the registry systems ORCID and ResearcherID are no alternatives to Scopus author ID, because a minority of laureates make use of these identifier systems and data is often incomplete. Unlike ORCID and ResearcherID that suffer from a selection bias as those scientists who remain in science maintain their author profiles, Scopus author ID exists for every author publishing in sources covered by Scopus [29].

Processing of bibliometric data is usually carried out using graph theory, such an application can be observed in the works Laudel G. (2005), Dyachenko E., Furukawa T.[30].

Summary

Using publications to collect bibliometric data (place, city or country of work) to analyze the migration of scientists is a good tool of data collecting. Such data allow us to analyze the mobility (double affiliation) or migration (loss of initial affiliation and the emergence of a new one) of scientists, as well as the dissemination of knowledge in a global and transnational space. The collection of bibliometric data seems possible in connection with the development of publications platforms. Foreign and leading Russian scientists can be analyzed by the publications presented in the SCOPUS and WoS databases. However, it is more advisable to use a domestic e-library for analyzing national migration and the earlier periods of time within Russia. In addition, the development of identification systems (ResearcherID and ORCID) implies a simplification of the procedure for identifying authors. However, as of 2015, this remains pending.

ACKNOWLEDGMENTS

Grant of the President of the Russian Federation "Migration of Russian scientists in cross-border and national space: problems, efficiency, assessment" (MK-1737.2019.6)

REFERENCES

- [1] Åkerblom, M., *Constructing internationally comparable indicators on the mobility of highly qualified workers*, (STI Review, Paris, 2001), 208 p.
- [2] OECD, *International mobility of the highly skilled*, (OECD Publications, Paris, 2001), 336 p.
- [3] Moguerou, P., Da Costa, O., Paola di Pietrogioacomo, M., & Laget, P., Indicators on researchers' career and mobility in Europe: A modeling approach. "Blue Skye II 2006": What indicators for science, technology and innovation policies in the 21st century? <http://www.oecd.org/sti/inno/37450342.pdf>.

- [4] Cañibano, C., Bozeman, B., "Curriculum vitae method in science policy and research evaluation: the state-of-the-art," *Res Eval* 18(2), 86–94 (2009).
- [5] Arvizu, D., Bowen, R., National Science Board. Science and Engineering Indicators 2014, National Science Foundation, Arlington, VA. <http://www.nsf.gov/statistics/seind14/>.
- [6] [OECD science, technology and industry scoreboard 2013: innovation for growth](#), OECD Publishing, Paris. doi:10.1787/sti_scoreboard-2013-en.
- [7] Sandstrum, U., "Combining curriculum vitae and bibliometric analysis: mobility, gender and research performance," *Res Eval* 18(2), 135–142 (2009).
- [8] Woolley, R., Turpin, T., "CV analysis as a complementary methodological approach: investigating the mobility of Australian scientists," *Res Eval* 18(2), 143–151 (2009).
- [9] Boring, P., Flanagan, K., Gagliardi, D., Kaloudis, A., Karakasidou, A., "International mobility: findings from a survey of researchers in the EU," *Sci Public Policy* 42(6), 811–826 (2015).
- [10] Flanagan, K., "International mobility of scientists," in *The handbook of global science, technology, and innovation*, edited by D. Archibugi, A. Filippetti (Wiley, Chichester, 2015), pp. 364–381.
- [11] De Filippo, D., Casado, E., Gymez, I., "Quantitative and qualitative approaches, to the study of mobility and scientific performance: a case study of a Spanish university," *Res Eval* 18(3), 191–200 (2009).
- [12] Dubois, P., Rochet, J., Schlenker, J., "Productivity and mobility in academic research: evidence from mathematicians," *Scientometrics* 98(3), 1669–1701 (2014).
- [13] Moed, H., Halevi, G., "A bibliometric approach to tracking international scientific migration," *Scientometrics* 101(3), 1–15 (2014).
- [14] Pierson, A., Cotgreave, P., "Citation figures suggest that the UK brain drain is a genuine problem," *Nature* 407(6800), 13–13 (2000).
- [15] Borjas, G., Doran, K., "The collapse of the Soviet Union and the productivity of American mathematicians," *Q J Econ* 127(3), 1143–1203 (2012).
- [16] Laudel, G., "Migration currents among the scientific elite," *Minerva* 43(4), 377–395 (2005).
- [17] Furukawa, T., Shirakawa, N., Okuwada, K., Sasaki, K., "International mobility of researchers in robotics, computer vision and electron devices: a quantitative and comparative analysis," *Scientometrics* 91(1), 185–202 (2012).
- [18] Deville, P., Wang, D., Sinatra, R., Song, C., Blondel, V., Barabasi, A., "Career on the move: geography, stratification, and scientific impact," *Sci Rep*, 4, 4770 (2014).
- [19] Laudel, G., "Studying the brain drain: Can bibliometric methods help?" *Scientometrics* 57(2), 215–237 (2003).
- [20] Laudel, G., "Migration currents among the scientific elite," *Minerva* 43(4), 377–395 (2005).
- [21] Zuckerman, H., *Scientific Elite – Nobel Laureates in the United States* (Transaction Publishers, New Brunswick, London, 1977), 34 p.
- [22] Hunter, R.S., Oswald, A.J., Charlton, B.G., "The elite brain drain," *The Economic Journal*, 119(538), F231–F251 (2009).
- [23] Dyachenko, E., "Internal migration of scientists in Russia and the USA: the case of physicists," *Scientometrics* 113(1), 105–122 (2017).
- [24] Trajtenberg, M., Shiff, G., Melamed, R., *The "Names Game": Harnessing Inventors' Patent Data for Economic Research* (NBER Working Paper, Cambridge, 2006), 75 p.
- [25] Agrawal, A., Cockburn, I., McHale, J., "Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships," *Journal of Economic Geography* 6(5), 571–591 (2006).
- [26] Dietz, J.S., Chompalov, I., Bozeman, B., O'Neil Lane, E., Park, J. "Using the curriculum vita to study the career paths of scientists and engineers: An exploratory assessment," *Scientometrics* 49, 419–442 (2000).
- [27] Cañibano, C., Otamendi, F.J., Solís, F., "International temporary mobility of researchers: a cross-discipline study," *Scientometrics* 89, 653 (2011).
- [28] Laudel, G., "Migration currents among the scientific elite," *Minerva* 43(4), 377–395 (2005).
- [29] Aman, V., "Does the Scopus author ID suffice to track scientific international mobility? A case study based on Leibniz laureates," *Scientometrics*, 117, 705–720 (2018).
- [30] Furukawa, T., Shirakawa, N., Okuwada, K., "Quantitative analysis of collaborative and mobility networks," *Scientometrics* 87(3), 451–466 (2011).